Chromosome 5

Jan-Fang Cheng, *Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

email: jfcheng@lbl.gov

A chromosome is a highly organized threadlike structure found in the nucleus of each human cell. Chromosomes are physical units of genetic materials and are made up largely by deoxyribonucleic acids and proteins. In humans, there are 23 pairs of chromosomes in a normal diploid cell. Chromosome 5 is the fifth chromosome primarily defined by its size.

Intermediate

Article contents

Human Chromosome Characteristics

0814.1 **Table 1** gives details of the characteristics of chromosome 5. (*See* A0061; A0270; A0272; A0279; A0325.)

Sources of the Sequence

0814.2 The first organized effort at mapping and sequencing this chromosome was launched in 1996 by the Human Genome Center of Lawrence Berkeley National Laboratory, which has become a member of the Joint Genome Institute (JGI). By early 2002, chromosome 5 in the GenBank consisted of over 2600 sequenced large insert clones, most of them derived from the Caltech D and the RPCI-11 bacterial artificial chromosome (BAC) libraries. Their initial positions and overlaps were determined by sequence-tagged site (STS) content, fluorescence *in situ* hybridization (FISH) and restriction maps, and were finally validated by sequence overlaps between clones. The tiling map of chromosome 5 (see Web links 5) has been continuously refined and used as a basis for the public sequence assembly. The assembled chromosome 5 sequence, the physical features and the annotation of gene structure can be accessed through at least three different Web browsers including the UCSC Genome Browser (Web links 2), the NCBI Map View (Web links 6) and the EMBL-EBI Ensembl Genome Browser (Web links 7). A finished version of this chromosome, not including the heterochromatic regions, is expected to be available by early 2003. (*See* A0346; A0353; A0355; A0346; A0353; A0355; A0356; A0779.)

Structural Features of the Chromosome

0814.3 Morton (1991) estimated the physical length of chromosome 5 to be 194 megabases (Mb). This estimate was based on the measurements of several autoradiography and flow cytometry experiments indicating that chromosome 5 represents 6.05% of the genome and that the total genome size is 3200 Mb. These numbers need to be confirmed by the finished sequence. (*See* A0007; A0783.)

Heterochromatic regions

0814.4 The centromere of chromosome 5 consists of at least three types of alphoid repeats. These repeats span approximately 3–4 Mb of deoxyribonucleic acid (DNA) arranged from p to q as D5Z12–D5Z2–D5Z1 (Puechberty *et al.*, 1999). Pulsed-field gel analysis revealed that the D5Z2 repeats alone represent 2–4 Mb of DNA, and the D5Z1 repeats are highly variable in copy number. Heterogeneity in copy number appears to be a characteristic of repetitive DNA in centromeres and telomeres. (*See* A0063; A0065; A0785; A0786; A0787.)

Euchromatic regions

0814.5 Over 88% of the 190 Mb euchromatic sequence has been determined and used in the analysis. Gene distribution of this chromosome is not random. Bands 5q31–q35 represent about 27% of the chromosome but contain 40% of the genes and 39% of the CpG islands. Interestingly, this region also contains a higher density of short interspersed nuclear elements (SINEs) (15.9%) and a lower density of long interspersed nuclear elements (LINEs) (17.0%) than the chromosome average (11.0% and 21.3% respectively). In a previous survey of 430 Mb of human DNA, a high density of Alu repeats was found in GC-rich or gene-rich regions, the opposite being true for LINE1 repeats. (*See* A0008; A0013; A0014; A0790.)

0814.T001 **Table 1** Characteristics of chromosome 5

Characteristic	Details
Chromosome number	5
Chromosome type	Metacentric
Physical size	
Overall	194 Mb
Short arm	50 Mb
Long arm	140 Mb
Genetic size[a]	
Overall average	197.5 cM
Male average	151.0 cM
Female average	244.8 cM
Average recombination length	1.06 cM/Mb
Physical characteristics	
Heterochromatic regions	
Distribution	Centromeric
Length	2–4 Mb
Repetitive sequence content[b]	
Nature	*Density*
Overall	45.66%
SINEs	11.02%
LINEs	21.33%
Variability[c]	
SNP density	793 SNPs/Mb
Microsatellite density	127 microsatellites/Mb (maximum)
Euchromatin[d]	
G+C content	39.9% (ranging from 30% to 62.2% per 20 kb)
CpG density	8.9 CpG islands/Mb
Genes[e]	
Gene density	8.2 genes/Mb
Gene number	1566
Confirmed	844
Predicted	722
Cytogenetic characteristics[f]	
Chromosome breakpoints	Rearrangements are relatively frequent; they include deletions, insertions, translocations and a ring chromosome
Disease related	Deletions and translocations associated with the *Cri du chat* syndrome on 5p, the 5q− syndrome, and spinal muscular atrophy on 5q
Cancer related	Recurrent balanced translocations with chromosomes 1, 2, 3, 7, 9, 12, 14, 17, 19 and 22, an inversion of p15 and q12, and recurrent unbalanced aberrations that occurred at 133 locations of chromosome 5 have been reported
Chromosome heteromorphism	A rare chromosome 5 heterochromatic variant 5qh that was derived from insertion of 9qh satellite 3 sequences was reported in a three-generation family

Mb: megabases (10^6 bases); cM: centimorgan; LINEs: long interspersed nuclear elements; SINEs: short interspersed nuclear elements; SNP: single nucleotide polymorphism.
[a]Marshfield (2001) (Web links 1).
[b] UCSC Genome Browser (2001) (Web links 2).
[c] NCBI Map View (2001) (Web links 3).
[d] UCSC Genome Browser (2001) (Web links 2).
[e] NCBI Map View (2001) (Web links 3).
[f]CGAP–Mitelman Database (Web links 4).

Recent duplication of a large deoxyribonucleic acid segment

0814.6 One interesting structural feature of chromosome 5 is that it contains at least one large duplication. The inverted duplication of about 500 kilobases (kb) containing the survival of motor neuron 1 telomeric (*SMN1*) and survival of motor neuron 2, centromeric (*SMN2*) genes was found in 5q13. The duplicated DNA is highly conserved, with an overall sequence divergence ranging from 0.15% to 0.34%, suggesting that the duplication is a recent evolutionary event. Rochette *et al.* (2001) compared the sequences of human and chimpanzee and suggested that the *SMN*

duplication occurred about 5–7 million years ago before the separation of chimpanzee and human. (*See* A0096; A0108; A1115; A1122; A1230.)

Interesting Gene Clusters and Biology

0814.7 There are many gene clusters on chromosome 5. Two cases are selected here to demonstrate the use of genomic structural (sequence and organization) information in approaching important biological questions.

Interleukin gene cluster on 5q31.1

0814.8 This region is of particular interest to immunologists because of the presence of a family of five hematopoietic growth factor genes (interleukin 3 (colony-stimulating factor, multiple) (*IL3*), interleukin 4 (*IL4*), interleukin 5 (colony-stimulating factor, eosinophil) (*IL5*), interleukin 13 (*IL13*) and colony-stimulating factor 2 (granulocyte-macrophage) (*CSF2*)) and a quantitative trait locus (QTL) associated with asthmatic airway inflammatory responses. The unique feature of this gene cluster is that although their gene products exert overlapping activities, these genes are not similar at nucleotide or amino acid level. It was hypothesized that these genes were physically bound together during evolution because they have been sharing common regulatory elements. To search for regulatory sequences, Loots *et al.* (2000) examined about 1 Mb of human and mouse DNA for conserved noncoding elements. One element, called CNS1, is 401 base pairs (bp) in size and is highly conserved (>80% identity) in several mammals. Characterization of this element in transgenic mice has revealed that CNS1 is a coordinate regulator of *IL4*, *IL13* and *IL5* in activated TH2 cells. This study has demonstrated the power of comparative sequence analysis in finding functional segments in noncoding genomic DNA. (*See* A0412; A1117; A1126.)

Protocadherin gene cluster on 5q31.3

0814.9 Protocadherins are thought to play a role in establishing and maintaining specificity in neuronal connections in the brain (Serafini, 1999). Analysis of protocadherin complementary DNA (cDNA) sequences has revealed that the 5′ portions of each cDNA differ from each other, whereas the 3′ portion of each cDNA is identical (Wu and Maniatis, 1999). These transcripts would encode a 'variable' extracellular and transmembrane domain and a 'constant' intracellular domain. Three closely linked protocadherin genes were identified in a 750 kb region of 5q31 through sequence search. Remarkably each gene contains a tandem array of 15–22 variable region exons and three constant region exons at the 3′ end. Such a gene structure is immediately reminiscent of that of the highly variable immunoglobulin (Ig) and T-cell-receptor genes, and this structure is believed to be involved in achieving diversity in B and T cells. The structural arrangement of these protocadherin genes has provided insights into a possible molecular basis for achieving neuronal diversity. (*See* A0045; A0126; A0145.)

Diseases Associated with Chromosome 5

Table 2 lists 94 diseases associated with chromosome 5. 0814.10 Affected genes were found in 73 diseases. Below we describe some recent developments in two disease loci that have a high incidence in the population. (*See* A1004; A1007.)

Spinal muscular atrophy

Spinal muscular atrophy (SMA) is a neuromuscular 0814.11 disease characterized by the degeneration of motor neurons in the spinal cord and brainstem nuclei, which leads to progressive weakness and wasting of the proximal muscles. SMA is the second most common lethal, autosomal recessive disease (after cystic fibrosis) in Caucasians, with an incidence of 1 in 10 000 and a carrier frequency of about 1 in 40. Three types of SMA have been characterized on the basis of clinical severity and age of onset. Type I is the most severe form and has the earliest onset; the affected individual usually dies before the age of two. Type II is the intermediate form with early onset and inability to walk. Type III is the mild form with an age of onset between 2 and 17 years. Mutations in the telomeric *SMN1* gene within the 500 kb inverted duplication in 5q13 are associated with the disease. Recent reports indicate the involvement of the SMN protein in ribonucleic acid (RNA) processing, particularly in the assembly of spliceosomal complexes and in the interaction with RNA polymerase II. (*See* A0036; A0323; A0513; A1011.)

Susceptibility to asthma

Chromosome 5q31–q33 has been implicated by several 0814.12 genetic linkage analyses to be a candidate locus for this quantitative trait. Among the analyzed traits, bronchial hyperresponsiveness (BHR) is a risk factor for asthma; the symptoms of BHR include a heightened bronchoconstrictor response to a variety of stimuli.

0814.T002 **Table 2** Diseases associated with human chromosome 5[a]

Chromosomal localization	Gene	Protein	OMIM	Disease
5p15.33	*SDHA*	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	600857	Leigh syndrome
5p15.31	*MTRR*	Methionine synthase reductase	602568	Homocystinuria–megaloblastic anemia, cbl E type
5p15.3	*SLC6A3*	Solute carrier family 6, member 3	126455	Attention deficit hyperactivity disorder
5p15.2	*ANKH*	A multipass transmembrane protein involved in the transport of pyrophosphate	123000	Craniometaphyseal dysplasia
5p15–p14	*DNAH5*	Dynein, axonemal heavy chain 5	603335	Primary ciliary dyskinesia
5p13.3	*NPR3*	Natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	108962	Hypertension, salt-resistant
5p13.2	*AMACR*	Alpha-methylacyl-CoA racemase	604489	Alpha-methylacyl-CoA racemase deficiency
5p13.2	*GDNF*	Glial cell derived neurotrophic factor precursor	600837	Hirschsprung disease
5p13	*OXCT*	Succinyl-CoA:3-oxoacid CoA transferase	245050	Ketoacidosis
5p13	*IL7R*	Interleukin-7 receptor, involved in the regulation of lymphopoiesis	600802	Severe combined immunodeficiency, T−B+NK+ type
5p12	*C6*	Complement component C6 precursor peptide	217050	C6 deficiency
5p12	*C6*	Complement component C6 precursor peptide	217050	Combined C6/C7 deficiency
5p12	*C7*	Complement component C7 precursor peptide	217070	C7 deficiency
5p12	*C9*	Complement component C9 precursor peptide	120940	C9 deficiency
5p12	*GHR*	Growth hormone receptor (serum binding protein), mutations in the extracellular domain	600946	Laron dwarfism
5p12	*GHR*	Growth hormone receptor (serum binding protein), mutations in the intracellular domain	245590	Laron dwarfism, type II, with elevated serum growth hormone binding protein
5p12	*GHR*	Growth hormone receptor (serum binding protein), abnormal GH receptor signaling	600946	Short stature, idiopathic, low serum growth hormone binding protein
5p	*MATP*	Membrane-associated transporter protein	606202	Oculocutaneous albinism, type IV
5q11	*MOCS2*	Molybdenum cofactor synthesis 2	603708	Molybdenum cofactor deficiency, type B
5q11.2	*NDUFS4*	NADH dehydrogenase (ubiquinone) Fe-S protein 4 (18kD) (NADH-coenzyme Q reductase)	602694	Complex I deficiency
5q11.2	*ITGA2*	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor) (collagen receptor)	192974	Glycoprotein Ia deficiency
5q11.2	*ITGA2*	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor) (collagen receptor)	192974	Neonatal alloimmune thrombocytopenia
5q11.2	*FST*	Follistatin, an activin-binding protein	184700	Polycystic ovary syndrome
5q11.2–q13.2	*DHFR*	Dihydrofolate reductase	126060	Anemia, megaloblastic, due to DHFR deficiency
5q12–q13	*MCCC2*	3-methylcotonyl-CoA carboxylase 2	210210	3-Methylcrotonylglycinuria
5q12.1	*CKN1*	Cockayne syndrome 1 (classical), a WD repeat protein	216400	Cockayne syndrome
5q12.2–q12.3	*DMGDH*	Dimethylglycine dehydrogenase	605849	Dimethylglycine dehydrogenase deficiency
5q13.2	*SMA1*	–	253300	Spinal muscular atrophy-1, or Werdnig–Hoffmann disease
5q13.2	*SMN1*	Survival motor neuron 1	600354	Spinal muscular atrophy-2
5q13.2	*SMN1*	Survival motor neuron 1	600354	Spinal muscular atrophy-3
5q13.3	*HEXB*	Hexosaminidase B (beta polypeptide)	268800	Sandhoff disease, infantile, juvenile and adult forms

0814.T002 **Table 2** Continued

Chromosomal localization	Gene	Protein	OMIM	Disease
5q13.3	*HEXB*	Hexosaminidase B (beta polypeptide)	268800	Spinal muscular atrophy, juvenile
5q14	*AP3B1*	Beta-3A-adaptin (probably involved in protein sorting in exocytic/endocytic pathways)	603401	Hermansky–Pudlak syndrome
5q14.1	*ARSB*	Arylsulfatase B (involved in degrading slow reacting substance of anaphylaxis)	253200	Maroteaux–Lamy syndrome
5q14.1	*MSH3*	DNA mismatch repair protein Msh3 (divergent upstream protein, DUP; mismatch repair protein 1, MRP1)	600887	Endometrial carcinoma
5q14.3	*RASA1*	RAS p21 protein activator (GTPase activating protein) 1	139150	Basal cell carcinoma
5q15–q21	*PCSK1*	Proprotein convertase subtilisin/kexin type 1	162150	Obesity with impaired prohormone processing
5q21	*MCC*	Mutated in colorectal cancers, encodes a 829 amino acid protein	159350	Colorectal cancer
5q22.2	*APC*	Adenomatous polyposis coli	175100	Adenomatous polyposis coli
5q22.2	*APC*	Adenomatous polyposis coli	175100	Colorectal cancer
5q22.2	*APC*	Adenomatous polyposis coli	175100	Desmoid disease, hereditary
5q22.2	*APC*	Adenomatous polyposis coli	175100	Gardner syndrome
5q22.2	*APC*	Adenomatous polyposis coli	175100	Turcot syndrome
5q23	*ADAMTS2*	Procollagen I N-proteinase (NPI)	225410	Ehlers–Danlos syndrome, type VII
5q23	*DTR*	Diphtheria toxin receptor	126150	Diphtheria, susceptibility to
5q23.1	*LOX*	Lysyl oxidase (protein-lysine 6-oxidase precursor)	153455	Cutis laxa, recessive, type I
5q23.1	*HSD17B4*	Hydroxysteroid (17-beta) dehydrogenase 4	601860	D-bifunctional protein deficiency
5q23.2	*FBN2*	Fibrillin 2 (congenital contractural arachnodactyly)	121050	Contractural arachnodactyly, congenital
5q31	*SLC22A5*	Sodium ion-dependent carnitine transporter	212140	Carnitine deficiency
5q31	*TTID*	Myotilin, a cytoskeletal protein, which directly interacts with α-actinin	159000	Limb girdle muscular dystrophy, type 1A
5q31	*FACL6*	Fatty acid CoA ligase, long chain 6	604443	Myelodysplastic syndrome
5q31	*GRAF*	GTPase regulator associated with focal adhesion kinase pp125	605370	Leukemia, juvenile myelomonocytic
5q31	*IL13*	Interleukin 13	147683	Asthma, susceptibility to
5q31–q32	*PDGFRB*	Platelet-derived growth factor receptor, beta polypeptide	173410	Myeloproliferative disorder with eosinophilia
5q31–q33	*SCA12*	An expansion of a CAG repeat in a brain-specific regulatory subunit of the protein phosphatase PP2A (PPP2R2B)	604326	Spinocerebellar ataxia 12
5q31–q34	*SCGB3A2*	Secretoglobulin, family 3A, member 2	606531	Asthma, susceptibility to
5q31.1	*IRF1*	Interferon regulatory factor-1	147575	Macrocytic anemia
5q31.1	*TGFBI*	Transforming growth factor induced protein	601692	Corneal dystrophy, Avellino type
5q31.1	*TGFBI*	Transforming growth factor induced protein	601692	Corneal dystrophy, Groenouw type I
5q31.1	*TGFBI*	Transforming growth factor induced protein	601692	Corneal dystrophy, lattice type I
5q31.1	*TGFBI*	Transforming growth factor induced protein	601692	Reis–Bucklers corneal dystrophy
5q31.1	*UBE2B*	Ubiquitin-conjugating enzyme E2B (RAD6 homolog) (ubiquitin carrier protein)	179095	Male infertility
5q31.1–q33.1	*GABRG2*	Gamma-aminobutyric acid A receptor, gamma-2	137160	Epilepsy, juvenile myoclonic

0814.T002 **Table 2** Continued

Chromosomal localization	Gene	Protein	OMIM	Disease
5q31.3	*DIAPH1*	Diaphanous homolog 1 (*Drosophila*)	602121	Deafness, autosomal dominant nonsyndromic sensorineural
5q31.3	*NR3C1*	Glucocorticoid receptor (GRL)	138040	Cortisol resistance
5q32	*ADRB2*	Adrenergic, beta-2-, receptor, surface	109690	Asthma, nocturnal, susceptibility to
5q32	*ADRB2*	Adrenergic, beta-2-, receptor, surface	109690	Obesity with impaired prohormone processing
5q32	*POU4F3*	POU domain, class 4, transcription factor 3	602460	Deafness, autosomal dominant
5q32	*SPINK5*	Serine protease inhibitor LEKTI	256500	Netherton syndrome
5q32	*SPINK1*	Serine protease inhibitor, Kazal type 1	167800	Pancreatitis, hereditary
5q32	*GLRA1*	Glycine receptor, alpha 1 (startle disease/hyperekplexia, stiff man syndrome)	138491	Startle disease, autosomal dominant and recessive
5q32–q33.1	*ACG1B*	A diastrophic dysplasia sulfate transporter	600972	Achondrogenesis, type IB
5q32–q33.1	*SLC26A2*	Solute carrier family 26 (sulfate transporter), member 2	256050	Atelosteogenesis, type II
5q32–q33.1	*SLC26A2*	Solute carrier family 26 (sulfate transporter), member 2	222600	Diastrophic dysplasia
5q32–q33.1	*SLC26A1*	Solute carrier family 26 (sulfate transporter), member 1	226900	Epiphyseal dysplasia, multiple
5q32–q33.1	*TCOF1*	Treacle protein, a putative nucleolar phosphoprotein	154500	Treacher–Collins–Franceschetti syndrome
5q33	*SGCD*	Sarcoglycan, delta, a 35 kDa dystrophin-associated glycoprotein	601287	Limb girdle muscular dystrophy, autosomal dominant, type 2F
5q33.1	*GM2A*	GM2 ganglioside activator protein (cerebroside sulfate activator protein, sphingolipid activator protein 3)	272750	GM2 gangliosidosis, AB variant
5q33.1	*CSF1R*	Colony stimulating factor 1 receptor, formerly McDonough feline sarcoma viral (v-fms) oncogene homolog	164770	Myeloid malignancy, predisposition to
5q33.1	*PDE6A*	Phosphodiesterase 6A, cGMP-specific, rod, alpha	180071	Retinitis pigmentosa, autosomal recessive
5q35	*NPM1*	Nucleophosphmin 1	164040	Leukemia, acute prommyelocytic
5q35	*NSD1*	Nuclear receptor-binding suvar, enhancer of zeste and trithorax domain protein 1	606681	Sotos syndrome
5q35.2	*NKX2E*	NK2 transcription factor homolog E (*Drosophila*)	600584	Atrial septal defect with atrioventricular conduction defects
5q35.2	*MSX2*	Msh homeo box homolog 2 (*Drosophila*)	123101	Craniosynostosis, type 2 (parietal foramina)
5q35.2	*F12*	Coagulation factor XII (Hageman factor)	234000	Factor XII deficiency
5q35.2	*PROP1*	Prophet of Pit1, paired-like homeodomain transcription factor	601538	Pituitary hormone deficiency, combined
5q35.2–q35.3	*B4GALT7*	Xylosylprotein 4-beta-galactosyltransferase polypeptide 7	604327	Ehlers–Danlos syndrome, progeroid form
5q35.3	*LTC4S*	Leukotriene C4 synthase	246530	Leukotriene C4 synthase deficiency
5q35.3	*FLT4*	Fms-related tyrosine kinase 4 (vascular endothelial growth factor receptor 3)	153100	Lymphedema, hereditary

[a]Sources of the disease information include the OMIM database (see Web links), the GeneCards Database (see Web links) and the Ensembl DiseaseView (see Web links).

This phenotype is associated with D5S436 in a linkage analysis of pairs of siblings, which is located near a major locus that governs serum IgE levels (Postma *et al.*, 1995). Transgenic mice of a panel of human yeast artificial chromosomes (YACs) spanning 1 Mb of 5q31 were used to screen for quantitative changes in several asthma-associated phenotypes (Symula *et al.*, 1999). A 180 kb region containing genes encoding for IL4 and IL13 was found to have altered the response of IgE to antigen treatment. This study demonstrates that altering gene dosage *in vivo* frequently affects quantitative traits normally influenced by that gene, and that functional screening offers a powerful tool for narrowing down the responsible genes. (*See* A0347; A0565; A0582.)

Complex Disorders Associated with Cytogenetic Abnormality

Cri du chat syndrome

0814.13 *Cri du chat* (CDC) syndrome appears to be one of the most common human deletion syndromes, with an incidence varying between 1 in 20 000 and 1 in 50 000 births or 1% of profoundly retarded patients (IQ less than 20) in the population. Although CDC syndrome has well-characterized clinical phenotypes, subjects with 5p deletion often show phenotypic and cytogenetic variability. Analyses of abnormal chromosomes from individuals with only a subset of the disease phenotypes have allowed sublocalization of genes responsible for specific disease traits. For examples, the speech delay was localized to the distal part of 5p15.3 (Church *et al.*, 1995); the cat-like cry was mapped to the proximal part of 5p15.3 between markers D5S731 and D5S760 (Overhauser *et al.*, 1994; Mainardi *et al.*, 2001); and one or two distinct regions for dysmorphism and mental retardation were assigned to 5p15.2 (Overhauser *et al.*, 1994; Church *et al.*, 1995; Mainardi *et al.*, 2001).

5q– syndrome and acute myelogenous leukemia

0814.14 Loss of all or part of the long arm of chromosome 5 is an anomaly that is frequently seen in patients with acute myelogenous leukemia (AML) and myelodysplasia. The common deletion of band 5q31 in malignant myeloid diseases signifies the existence of a key negative regulator of leukemogenesis. Physical mapping has narrowed the common deleted region for a putative tumor suppressor gene to a 1.5 Mb interval between D5S479 and D5S500 (Zhao *et al.*, 1997). The 5q– syndrome is a distinct form of myelodysplastic syndrome, in which the erythroid and megakaryocytic lineages are predominantly affected. FISH and gene dosage analysis have mapped the critical region of gene loss within 5q32 which is distal to the AML locus (Boultwood *et al.*, 2000).

References

Boultwood J, Strickson AJ, Jabs EW, *et al.* (2000) Physical mapping of the human ATX1 homologue (HAH1) to the critical region of the 5q-syndrome within 5q32, and immediately adjacent to the SPARC gene. *Human Genetics* **106**: 127–129.

Church DM, Bengtsson U, Nielsen KV, Wasmuth JJ and Niebuhr E (1995) Molecular definition of deletions of different segments of distal 5p that result in distinct phenotypic features. *American Journal of Human Genetics* **56**: 1162–1172.

Loots GG, Locksley RM, Blankespoor CM, *et al.* (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**: 136–140.

Mainardi PC, Perfumo C, Cali A, *et al.* (2001) Clinical and molecular characterisation of 80 patients with 5p deletion: genotype–phenotype correlation. *Journal of Medical Genetics* **38**: 151–158.

Morton NE (1991) Parameters of the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 7474–7476.

Overhauser J, Huang X, Gersh M, *et al.* (1994) Molecular and phenotypic mapping of the short arm of chromosome 5: sublocalization of the critical region for the *cri-du-chat* syndrome. *Human Molecular Genetics* **3**: 247–252.

Postma DS, Bleecker ER, Amelung PJ, *et al.* (1995) Genetic susceptibility to asthma – bronchial hyperresponsiveness coinherited with a major gene for atopy. *New England Journal of Medicine* **333**: 894–900.

Puechberty J, Laurent AM, Gimenez S, *et al.* (1999) Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen. *Genomics* **56**: 274–287.

Rochette CF, Gilbert N and Simard LR (2001) SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to *Homo sapiens*. *Human Genetics* **108**: 255–266.

Serafini T (1999) Finding a partner in a crowd: neuronal diversity and synaptogenesis. *Cell* **98**: 133–136.

Symula DJ, Frazer KA, Ueda Y, *et al.* (1999) Functional screening of an asthma QTL in YAC transgenic mice. *Nature Genetics* **23**: 241–244.

Wu Q and Maniatis T (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**: 779–790.

Zhao N, Stoffel A, Wang PW, *et al.* (1997) Molecular delineation of the smallest commonly deleted region of chromosome 5 in malignant myeloid diseases to 1–1.5 Mb and preparation of a PAC-based physical map. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 6948–6953.

Further Reading

Fairman J, Chumakov I, Chinault AC, Nowell PC and Nagarajan L (1995) Physical mapping of the minimal region of loss in 5q- chromosome. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 7406–7410.

Fischer U, Liu Q and Dreyfuss G (1997) The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* **90**: 1023–1029.

Gu Z, Wang H, Nekrutenko A and Li WH (2000) Densities, length proportions, and other distributional features of repetitive

sequences in the human genome estimated from 430 megabases of genomic sequence. *Gene* **259**: 81–88.
Horrigan SK, Arbieva ZH, Xie HY, *et al.* (2000) Delineation of a minimal interval and identification of 9 candidates for a tumor suppressor gene in malignant myeloid disorders on 5q31. *Blood* **95**: 2372–2377.
Liu Q, Fischer U, Wang F and Dreyfuss G (1997) The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* **90**: 1013–1021.
Niebuhr E (1978) The *cri du chat* syndrome: epidemiology, cytogenetics, and clinical features. *Human Genetics* **44**: 227–275.
Pellizzoni L, Charroux B, Rappsilber J, Mann M and Dreyfuss G (2001) A functional interaction between the survival motor neuron complex and RNA polymerase II. *Journal of Cell Biology* **152**: 75–85.

Web links

1. Marshfield Clinic/Center for Medical Genetics
   http://research.marshfieldclinic.org/genetics
2. UCSC Genome Bioinformatics Genome Browser
   http://genome.ucsc.edu
3. NCBI Entrez Map View
   http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/maps.cgi?org=hum&chr=5
4. CGAP–Mitelman Database
   http://cgap.nci.nih.gov/Chromosomes/Mitelman
5. DOE Joint Genome Institute. Tiling map of chromosome 5
   http://www.jgi.doe.gov/programs/hgp.html
6. NCBI Entrez Genome Map View
   http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map_search
7. EMBL-EBI Ensembl Human Genome Browser
   http://www.ensembl.org/Homo_sapiens
8. Online Mendelian Inheritance in Man (OMIM) Database
   http://www3.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
9. GeneCards Database
   http://nciarray.nci.nih.gov/cgi-bin/cards/listdiseasecards?search=5&type=chrom
10. Ensembl DiseaseView
    http://www.ensembl.org/Homo_sapiens/diseaseview

Survival of motor neuron 1 telomeric (*SMN1*); Locus ID: 6606 LocusLink:
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=6606

Survival of motor neuron 1 telomeric (*SMN1*); MIM number: 600354 OMIM:
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?600354

Glossary

Quantitative trait loci (QTL): The location of a gene that affects a phenotypic trait that is measured on a quantitative scale. These traits are typically affected by more than one gene, and sometimes by the environment.

Interleukins (IL): A family of glycoproteins secreted by a variety of leukocytes that have effects on other leukocytes. The effects are mediated through specific receptors on the surface of target cells, which are coupled to intracellular signal transduction and second messenger pathways to activate cell growth and/or differentiation. The name 'interleukin' came from 'between leukocytes'.

Myelodysplasia (MDS): A group of disorders in which the bone marrow overproduces cells, but they do not mature normally. Most MDS patients are anemic and many have low number of neutrophils (infection fighting white blood cells) and low platelet counts. MDS is a chronic disorder, but it usually evolves to acute myelogenous leukemia (AML) over time. The standard treatment for MDS is supportive care, transfusions and antibiotics as needed. The only curative therapy for MDS is allogeneic bone marrow transplantation.

Satellite sequence: DNA with very high repetitions (from hundreds to over 10 000 copies) of a basic motif or repeat unit (commonly 100–300 base pairs), which mostly occur at centromeres and telomeres of eucaryotic chromosomes.

Draft sequence: An unfinished version of genomic sequence that is produced by assembling overlapping sequence reads from individual clones or from a whole genome. The quality of a draft sequence is usually defined by the redundancy of the high-quality sequenced nucleotides representing the target.

BAC library: A genomic DNA library containing large DNA inserts, typically 100–200 kb in size. The vector DNA contains all the necessary elements to maintain a single copy per cell in the bacterial cells; therefore, this type of library is called a bacterial artificial chromosome (or BAC) library.

Keywords

chromosome 5, complex genetic diseases, spinal muscular atrophy, asthma, interleukin gene cluster, Cri du chat, acute myelogenous leukemia, protocadherin